**Montserrat HERNANDEZ-LOPEZ, PhD**
**E-mail: mhdezl@ull.edu.es**
**José Juan CACERES-HERNANDEZ, PhD**
**E-mail: jcaceres@ull.es**
**University of La Laguna**

# FORECASTING THE COMPOSITION OF SPAIN'S UNEMPLOYED POPULATION BY GENETIC ALGORITHMS

***Abstract.****A genetic algorithm is developed to forecast the relative presence of different profiles in Spain's unemployed population. A selection operator is defined that assumes that the higher the unemployment rate of a profile, the higher the probability that such a profile is present in future populations. A transition matrix takes other factors into account which may influence changes in the profiles. The algorithm is applied to the original quarterly populations of Spain's unemployed in 2014. Then, it is applied to obtain the forecast of the quarterly populations of unemployed in Spain in 2015 and 2016. This methodological proposal is shown to provide the type of forecast that is very useful in policy-making decisions to reduce the higher unemployment rates caused by the economic crisis in the euro area.*

***Keywords****: unemployment, forecasting, genetic algorithms, transition matrix.*

**JEL Classification: C63, J64**

## 1. Introduction

The Spanish economic crisis has led to higher levels of unemployment. In fact, according to the Labour Force Survey of the Spanish National Institute of Statistics, the number of unemployed exceeded 6 million people in the first quarter of 2013, which meant an unemployment rate of over 27%. Although unemployment has fallen below 5 million in the last couple of years, and the unemployment rate was under 19% in the last quarter of 2016, there are still significant disparities in this rate across regions and demographic groups according to gender, age, education level or migrant condition.

There are differences in the dynamics of unemployment rates between men and women, and between young people and adults, as pointed out by Altuzarra (2015). Regarding gender, the results obtained by Pastor et al. (2016) reveal that the gender gap decreases as the educational level increases. In addition, according to Antón et al. (2012), migrant condition also implies different access to employment.

The effects of the Spanish economic crisis on the unemployment rates by gender, age, education level or nationality have been analyzed by Motellón and



DOI: 10.24818/18423264/52.1.18.10

López-Bazo (2014), Peña-Boquete (2014), Gil-Alonso and Vidal-Coso (2015), and Sacht (2015), among others. In this sense, Spain is no exception because a gender unemployment gap has been also highlighted for different European countries.

Indeed, a number of studies have highlighted a range of factors that affect unemployment rates. According to Li and Zhang (2010), the Chinese job-market prefers male graduates over their female peers with similar qualifications. The relationship between age and unemployment rate is dealt with in Caliendo and Schmidl (2016) or Marelli and Vakulenko (2016). According to Klein (2015), higher education reduces the incidence of unemployment. Following Uhlendorff and Zimmermann (2014), unemployment rates are often higher for migrants than for natives.

Furthermore, the dynamics of unemployment vary across different demographic groups. Peiró et al. (2012) found that the association between unemployment and the business cycle is more intense for males than for females. Perugini and Signorelli (2010) showed the stability of youth unemployment rates at high levels in European regions, whereas Ponomareva and Sheen (2013) explained that younger people have a disproportionate difficulty in finding and keeping jobs in business cycle downturns. Ghosray et al. (2016) also pointed out that youth unemployment is more sensitive to business cycles. Dustmann et al. (2010) emphasize that economic shocks bring about significantly larger unemployment for low-skilled workers relative to high-skilled workers and for immigrants relative to natives within the same skill group.

Therefore, as suggested by Queneau and Sen (2012), policy makers should be aware of both the level of unemployment rates within each demographic group as well as the varying levels of its persistence among these groups. From this point of view, forecasting procedures are useful tools to design public policy. Many papers have dealt with forecasting the aggregate level of unemployment or the unemployment rate by using time series analysis, neural networks and genetic algorithms or alternative approaches. As regards the Spanish case, García-Cintado et al. (2014), Altuzarra (2015) and Cuestas and Gil-Alana (2017) have analyzed the statistical properties of the unemployment rate over several decades starting from 1976. Interesting proposals to forecast the aggregate unemployment rate have been made by Olmedo (2014) and Vicente et al. (2015). However, an aggregate forecast is not enough to take into account the different dynamics of unemployment across demographic groups. Statistical procedures capable of forecasting the relative weight of each of these demographic groups in the total unemployed population is also needed. The aim of this paper is to forecast, by means of a genetic algorithm, the composition of Spain's unemployed people as a function of their individual characteristics.

In the following section, the methodological elements of the genetic algorithm are presented. Next, the results obtained by applying the proposed algorithm are shown. Finally, concluding remarks are stated and some areas that this research has implications for are indicated.



DOI: 10.24818/18423264/52.1.18.10

## 2. Material and methods

The statistical tool proposed in this paper to forecast the composition of the population of unemployed people is, as mentioned, a genetic algorithm. Some applications in economics are mentioned in Santana and Coello (2006), Waltman and van Eck (2012) and Marqués et al. (2013), among others. In some of them, genetic algorithms were applied as a statistical procedure to drive an optimization process. However, in this paper, a genetic algorithm, based on the natural selection principle that drives transformations in different species populations, is applied as a mechanism that describes the dynamic process of transformation in the population of unemployed people in Spain. First, the statistical procedure is developed. Then, the active population from which the operators in the genetic algorithm are estimated and the original populations to which the genetic algorithm is applied are described.

### 2.1. The design of the genetic algorithm

The algorithm modifies the original population of unemployed people in quarter $k$of year $T$, and final populations of unemployed people in the same quarter of years $T+1$ or $T+2$ are predicted. Individuals in these populations are identified by a structure or a chain that indicates their characteristics. The transformations that determine the relative frequency corresponding to each structure in the final population is guided by the selection operator in the first of two phases, and, once an intermediate population is obtained, a transition matrix operates in the second phase to transform the intermediate population into the final population.

#### 2.1.1. Fitness function and selection operator

The members of the active population whose unemployment rate is higher are expected to increase their relative frequency in future populations of unemployed people. This is the basic assumption to define the selection operator.

The so called fitness function is defined from the unemployment rates for individual profiles in terms of gender, age, education level and nationality. Suppose that, given the categories of the individual characteristics,$m$ different structures, $\{E_i\}_{i=1,\dots,m}$, are defined. If the composition of quarterly populations of active people in quarter $k$until year $T$is observed, then, to forecast the population of unemployed people in quarter $k$ of year $T+h$, the fitness function for an individual whose characteristics correspond to structure $E_i$, $i=1,\dots,m$, is defined as the forecast of the unemployment proportion among active individuals with such a structure. These forecasts are obtained from the unemployment proportions by profile observed in the past. Let $f_{i,k,t}$, $i=1,\dots,m$, be the unemployment proportion among the active individuals with structure $E_i$in quarter $k$ of year $t$, defined as

$$f_{i,k,t}=\frac{n^u_{i,k,t}}{n^a_{i,k,t}}, \qquad (1)$$



DOI: 10.24818/18423264/52.1.18.10

where $n^{u}_{i,k,t}$ is the number of unemployed people with structure $E_i$in the active population in quarter $k$ of year $t$, whereas $n^{a}_{i,k,t}$ is the total number of individuals with structure $E_i$in the active population at the same point in time.

Thus, from the series of unemployment proportions observed for structure $E_i$in quarter $k$,$\{f_{i,k,t}\}_{t=1,\dots,T}$, the three-segment cubic spline

$$f_{i,k,t} = \sum_{\tau=1}^{3}\left(\alpha^{u,c}_{i,k,\tau} + \beta^{u,c}_{i,k,\tau}t + \gamma^{u,c}_{i,k,\tau}t^2 + \delta^{u,c}_{i,k,\tau}t^3\right)D^{u,c}_{i,k,\tau} + \varepsilon^{u,c}_{i,k,t}, t = 1, \dots, T, \quad (2.a)$$

where

$$D^{u,c}_{i,k,1} = \begin{cases} 1 & , \quad 1 \le t < t^{u,c}_1 \\ 0 & , \quad in\ other\ case \end{cases}$$
$$D^{u,c}_{i,k,2} = \begin{cases} 1 & , \quad t^{u,c}_1 \le t < t^{u,c}_2 \\ 0 & , \quad in\ other\ case \end{cases}$$
$$D^{u,c}_{i,k,3} = \begin{cases} 1 & , \quad t \ge t^{u,c}_2 \\ 0 & , \quad in\ other\ case \end{cases}$$

and also the three-segment quadratic spline

$$f_{i,k,t} = \sum_{\tau=1}^{3}\left(\alpha^{u,q}_{i,k,\tau} + \beta^{u,q}_{i,k,\tau}t + \gamma^{u,q}_{i,k,\tau}t^2\right)D^{u,q}_{i,k,\tau} + \varepsilon^{u,q}_{i,k,t}, t = 1, \dots, T, \quad (2.b)$$

where

$$D^{u,q}_{i,k,1} = \begin{cases} 1 & , \quad 1 \le t < t^{u,q}_1 \\ 0 & , \quad in\ other\ case \end{cases}$$
$$D^{u,q}_{i,k,2} = \begin{cases} 1 & , \quad t^{u,q}_1 \le t < t^{u,q}_2 \\ 0 & , \quad in\ other\ case \end{cases}$$
$$D^{u,q}_{i,k,3} = \begin{cases} 1 & , \quad t \ge t^{u,q}_2 \\ 0 & , \quad in\ other\ case \end{cases}$$

are estimated by ordinary least squares. The cubic spline is forced to satisfy continuity constraints of the function and its first and second derivatives in the break points$t^{u,c}_1$and $t^{u,c}_2$, whereas the quadratic spline is forced to satisfy continuity constraints of the function and its first derivatives in the corresponding break points $t^{u,q}_1$ and $t^{u,q}_2$ (see Poirier, 1976). The locations of the break points are selected in order to capture the main turning points in the values observed until year $T$ and improve the forecasting performance.

The forecasts of unemployment rates by profile in quarter $k$ of year $T+1$ are obtained from the estimates of the cubic spline model as

$$\hat{f}_{i,k,T+1} = \hat{\alpha}^{u,c}_{i,k,3} + \hat{\beta}^{u,c}_{i,k,3}(T+1) + \hat{\gamma}^{u,c}_{i,k,3}(T+1)^2 + \hat{\delta}^{u,c}_{i,k,3}(T+1)^3. \quad (3.a)$$

On the other hand, the forecasts of unemployment rates by profile in quarter $k$ of year $T+2$ are obtained from the estimates of the quadratic spline model by

$$\hat{f}_{i,k,T+2} = \hat{\alpha}^{u,q}_{i,k,3} + \hat{\beta}^{u,q}_{i,k,3}(T+2) + \hat{\gamma}^{u,q}_{i,k,3}(T+2)^2. \quad (3.b)$$

Note that in a cubic spline, the curvature appears to change in a smoother way than in a quadratic spline, but the latter prevents an excessively large change in the level for a two-year forecast.

As the first step in the execution of the genetic algorithm, such fitness functions are applied to the population of unemployed people observed in quarter $k$



DOI: 10.24818/18423264/52.1.18.10

of year $T$. Once the structures observed in the original population are ordered according to individual characteristics, a selection operator is applied to obtain an intermediate population in quarter $k$ ofyear $T+h$, $h=1,2$ in such a way that the probability that an individual is selected is proportional to the value of the fitness function for such an individual. That is to say, the probability to select an individual with structure $E_i$ in quarter $k$ of year $T+h$ is

$$p^{s}_{i,k(T+h)}=\frac{\hat{f}_{i,k,T+h}\hat{n}^{a}_{i,k,T+h}}{\sum_{i=1,\dots,m}\hat{f}_{i,k,T+h}\hat{n}^{a}_{i,k,T+h}}, \quad (4)$$

where$\hat{n}^{a}_{i,k,T+h}$ is the forecast of the total number of individuals with structure $E_i$ in the active population in quarter $k$ of year $T+h$. Such forecast for structure $E_i$ in quarter $k$ of year $T+1$ is obtained from the series $\{n^{a}_{i,k,t}\}_{t=1,\dots,T}$, by estimating the three-segment cubic spline

$$n^{a}_{i,k,t}=\sum_{\tau=1}^{3}\left(\alpha^{a,c}_{i,k,\tau}+\beta^{a,c}_{i,k,\tau}t+\gamma^{a,c}_{i,k,\tau}t^2+\delta^{a,c}_{i,k,\tau}t^3\right)D^{a,c}_{i,k,\tau}+\varepsilon^{a,c}_{i,k,t}, t=1,\dots,T, \quad (5.a)$$

where

$$D^{a,c}_{i,k,1}=\begin{cases}1 & , \quad 1\le t<t^{a,c}_{1}\\ 0 & , \quad in\ other\ case\end{cases}$$
$$D^{a,c}_{i,k,2}=\begin{cases}1 & , \quad t^{a,c}_{1}\le t<t^{a,c}_{2}\\ 0 & , \quad in\ other\ case\end{cases}$$
$$D^{a,c}_{i,k,3}=\begin{cases}1 & , \quad t\ge t^{a,c}_{2}\\ 0 & , \quad in\ other\ case\end{cases}$$

and usual continuity constraints are applied. To obtain the forecast for year $T+2$, the three-segment quadratic spline

$$n^{a}_{i,k,t}=\sum_{\tau=1}^{3}\left(\alpha^{a,q}_{i,k,\tau}+\beta^{a,q}_{i,k,\tau}t+\gamma^{a,q}_{i,k,\tau}t^2\right)D^{a,q}_{i,k,\tau}+\varepsilon^{a,q}_{i,k,t}, t=1,\dots,T, \quad (5.b)$$

where

$$D^{a,q}_{i,k,1}=\begin{cases}1 & , \quad 1\le t<t^{a,q}_{1}\\ 0 & , \quad in\ other\ case\end{cases}$$
$$D^{a,q}_{i,k,2}=\begin{cases}1 & , \quad t^{a,q}_{1}\le t<t^{a,q}_{2}\\ 0 & , \quad in\ other\ case\end{cases}$$
$$D^{a,q}_{i,k,3}=\begin{cases}1 & , \quad t\ge t^{a,q}_{2}\\ 0 & , \quad in\ other\ case\end{cases}$$

is proposed to be estimated once the continuity constraints are imposed. The forecasts of the number of people in the active population by profile in quarter $k$ of year $T+1$ are obtained from the estimates of the cubic spline model as

$$\hat{n}^{a}_{i,k,T+1}=\hat{\alpha}^{a,c}_{i,k,3}+\hat{\beta}^{a,c}_{i,k,3}(T+1)+\hat{\gamma}^{a,c}_{i,k,3}(T+1)^2+\hat{\delta}^{a,c}_{i,k,3}(T+1)^3. \quad (6.a)$$

On the other hand, the corresponding forecasts for quarter $k$ of year $T+2$ are obtained from the estimates of the quadratic spline model as

$$\hat{n}^{a}_{i,k,T+2}=\hat{\alpha}^{a,q}_{i,k,3}+\hat{\beta}^{a,q}_{i,k,3}(T+2)+\hat{\gamma}^{a,q}_{i,k,3}(T+2)^2. \quad (6.b)$$

In formal terms, to obtain an intermediate population of size $n$, the selection operator applies the following procedure. Let $I_{1,i}$ be an individual with structure $E_i$ in the original population, in such a way that $\Omega_1\colon\{I_{1,1},\dots,I_{1,m}\}$ denotes the set of $m$



DOI: 10.24818/18423264/52.1.18.10

different individuals in the original population, and let $W: \{1, \dots, n\}$ be defined as the set of $n$ positions where the individuals copied from the original population are located, in such a way that the characteristics of any individual copied at the selection phase correspond to one of the structures $E_i$, $i = 1, \dots, m$. Thus, the intermediate population $\Omega_2: \{I_{2,1}, \dots, I_{2,n}\}$ is obtained through the selection operator $s(j) = I_{2,j}$, and is defined as $s: W \longrightarrow \Omega_1$, in such a way that

$$P\left(s(j) = I_{1,i}\right) = P\left(I_{2,j} = I_{1,i}\right) = p^s_{i,k(T+h)},\ i = 1, \dots, m, \forall j = 1, \dots, n \quad (7)$$

where the probability that individual $i$ in the original population is copied in position $j$ in the intermediate population, $p^s_{i,k(T+h)}$, is defined as in Equation (4). The intermediate population is obtained by randomly generating the results of $n$ multinomial experiments of size $m$ with probabilities $p^s_{1,k(T+h)}, \dots, p^s_{m,k(T+h)}$.

### 2.1.2. Transition matrix

In the second phase, an element of heterogeneity should be incorporated to allow an individual whose selection probability is not so high. The crossover and mutation operators are normally responsible for doing this because they allow the characteristics that identify the selected individual in the first phase to be modified. These operators are the elements of a simple genetic algorithm, but, as Hernández-López and Cáceres-Hernández (2007)pointed out, the transformations in the population that they produce are completely random. Nevertheless, following the proposal in Hernández-López and Cáceres-Hernández (2016), in this paper some transformations are assigned a greater likelihood than others. In this sense, once the copies have been selected, a transition matrix is defined to assign specific probabilities to each of the transformations. In other words, in a second phase, each of the structures copied in the intermediate population remains in the final population or is transformed into another structure according to the probabilities in the corresponding row of the transition matrix as indicated.

By means of a square transition matrix $M$, each structure $E_i$, $i = 1, \dots, m$, from the intermediate population is assumed to be transformed into another structure $E_j$ from the final population in quarter $k$ of year $T + h$ with the probabilities $p^{Tr}_{i,j,k(T+h)}, j = 1, \dots, m$, located at the $i^{\text{th}}$ row of the matrix $M$.

The formal definition of the transition matrix is as follows. Let $\Omega_2: \{I_{2,1}, \dots, I_{2,n}\}$ be a set of $n$ individuals from the resultant intermediate population from the copies, $E$ a set of $m$ structures in which each individual in $\Omega_2$ can be transformed, and $R: \{1, \dots, n\}$ a set of $n$ positions where individuals from the final population can be placed. This final population $\Omega_3: \{I_{3,1}, \dots, I_{3,n}\}$ is obtained using the transition matrix operator $tm(r) = I_{3,r}$, defined as $tm(q): R \longrightarrow E$, such that

$$P\left(tm(r) = E_j\right) = P\left(I_{3,r} = E_j\right) = p^{Tr}_{r,j,k(T+h)},\ r = 1, \dots, m, j = 1, \dots, m, \quad (8)$$

is the probability that the individual located at position $r$ in the intermediate population, $I_{2,r}$, is transformed into another individual in the final population



DOI: 10.24818/18423264/52.1.18.10

whose structure is $E_j$. If the individual that occupies position $r$in the intermediate population in quarter$k$ of year $T+h$has structure $E_i$, then $p^{Tr}_{r,j,k(T+h)} = p^{Tr}_{i,j,k(T+h)}$, that is, the element in column$j$ of row$i$ in the transition matrix. In this way, a multinomial experiment of size $m$ with probabilities$p^{Tr}_{i,1,k(T+h)}, \dots, p^{Tr}_{i,m,k(T+h)}$ can be performed to determine which individual $I_{3,r}$is transformed from individual $I_{2,r}$.

In this paper, the probabilities in the transition matrix, $M$ , are defined as

$$p^{Tr}_{i,j,k(T+h)} = \begin{cases} \alpha & , \quad i=j \\ (1-\alpha)\dfrac{1-|\hat{f}_{i,k,T+h}-\hat{f}_{j,k,T+h}|}{\sum_{\substack{j=1,\dots,m \\ j\neq i}} 1-|\hat{f}_{i,k,T+h}-\hat{f}_{j,k,T+h}|} & , \quad j \neq i \end{cases} \quad (9)$$

in such a way that the shorter the distance between the forecasts of the unemployment rates corresponding to structures $E_i$ and $E_j$, the higher their transition probability. If structures $E_i$ and $E_j$ are equal, then $p^{Tr}_{r,j,k(T+h)} = \alpha$, $i,j = 1,\dots,m$. Therefore, $\sum_{j=1,\dots,m} p^{Tr}_{r,j,k(T+h)} = 1$, $\forall i = 1,\dots,m$.

Note that the transition matrix is useful to introduce random changes that are not driven by the same principle as the one that drives the selection operator. Therefore, when the changes observed in the original population are not expected according to the selection operator, this matrix can improve the performance of the genetic algorithm to forecast the composition of a population.

### 2.2. Data sources

Before applying the genetic algorithm designed in the previous section, the sources of data need to be explained. Individual data about the Spanish active population from the first quarter in 2005 to the fourth quarter in 2014, provided by INE (Spanish National Statistics Institute) in the Active Population Surveys corresponding to the mentioned period, were used to obtain the forecast of the unemployment rates by structure, $\hat{f}_{i,k,T+h}$, in the four quarters in 2015 and 2016. The unemployed populations in each one of the four quarters in 2014 were defined as the original populations to which the genetic algorithm is applied to forecast the composition of these populations in each of the four quarters in 2015 and 2016, respectively.

Spanish Active Population Surveys provide information about the labour status of people over 15 years old. Once the economically inactive population is eliminated, for each individual in the active population in a quarter, characteristics and categories are defined as shown in Table1. The quarterly series of unemployment rates by these categories are shown in Figure 1. The values of the unemployment rates, $f_{i,k,t}$, corresponding to each one of the 96 different structures which can be defined from categories of individual characteristics $X(G)$,$X(A)$,$X(E)$, and $X(N)$, are not shown to save space. However, Figure 2 shows the changes observed in these rates over time for the more frequent structures in the active population in Spain, whereas in Figure 3 the relative frequencies of these structures are shown. Figure 4 shows the relative frequencies



DOI: 10.24818/18423264/52.1.18.10

for the categories of individual characteristics in the active population in the sample.

**Table 1. Individual characteristics and categories**

| Characteristic | Categories | Characteristic | Categories |
|---|---|---|---|
| $Y$: Labour status | 1 (unemployed)<br>2 (employee) | $X(G)$: Gender | 1 (male)<br>2 (female) |
| $X(A)$: Age | 1 (16-24 years old)<br>2 (25-34 years old)<br>3 (35-44 years old)<br>4 (45-54 years old)<br>5 (55-64 years old)<br>6 (65 years old and over) | $X(E)$: Education level | 0 (no education)<br>1 (primary school)<br>2 (secondary studies)<br>3 (university studies) |
| | | $X(N)$: Nationality | 1 (Spanish)<br>2 (Non-Spanish) |

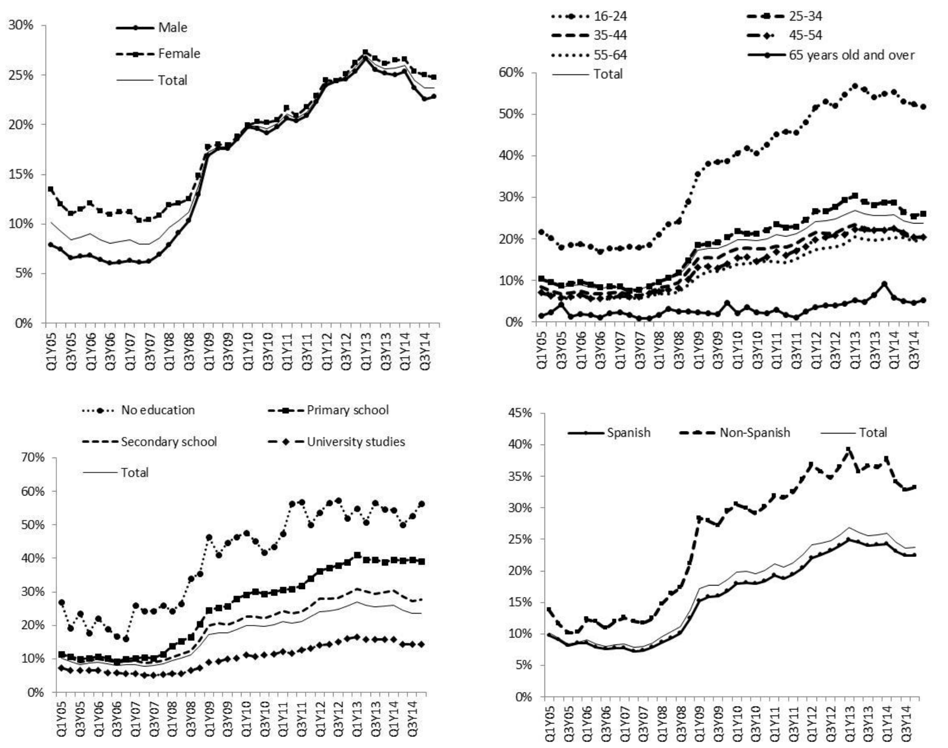


**Figure 1. Unemployment rates for categories of individual characteristics in the Spanish active population**



DOI: 10.24818/18423264/52.1.18.10

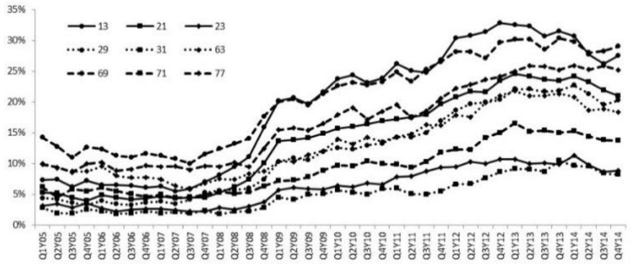


**Figure 2. Unemployment rates for the most frequent structures in the Spanish active population**

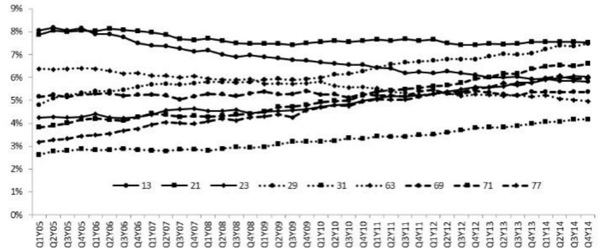


**Figure 3. Relative frequencies for the most frequent structures in the Spanish active population**

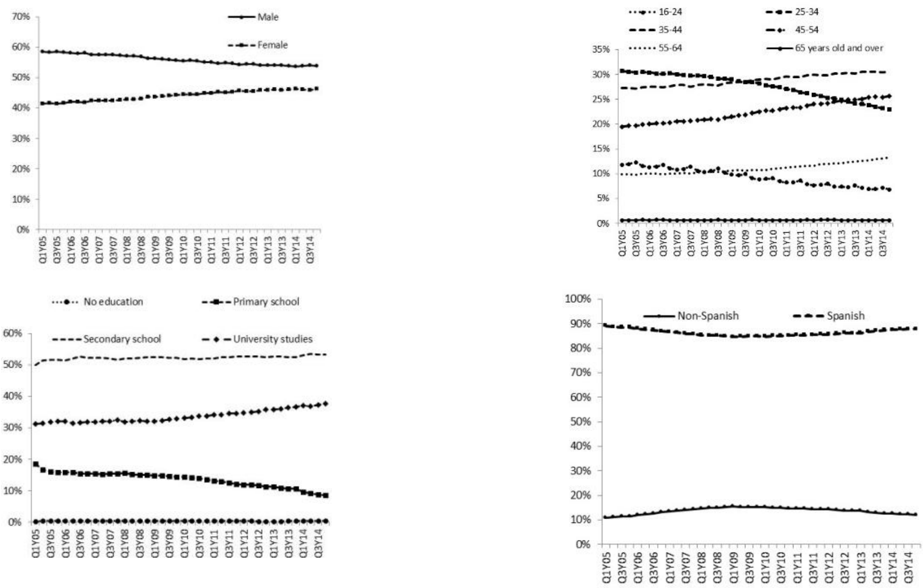


**Figure 4. Relative frequencies for categories of individual characteristics in the Spanish active population**



DOI: 10.24818/18423264/52.1.18.10

In a similar sense, Figure 5 shows the changes observed in the relative frequencies for the most frequent structures in the Spanish unemployed population, whereas the relative frequencies corresponding to categories of the individual characteristics are shown in Figure 6.

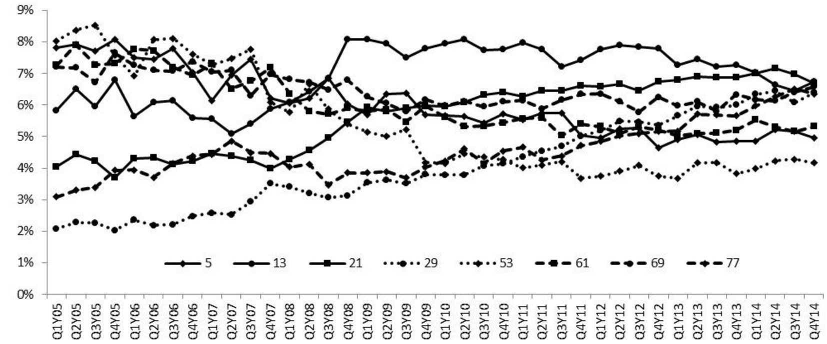


**Figure 5. Relative frequencies for the most frequent structures in the Spanish unemployed population**

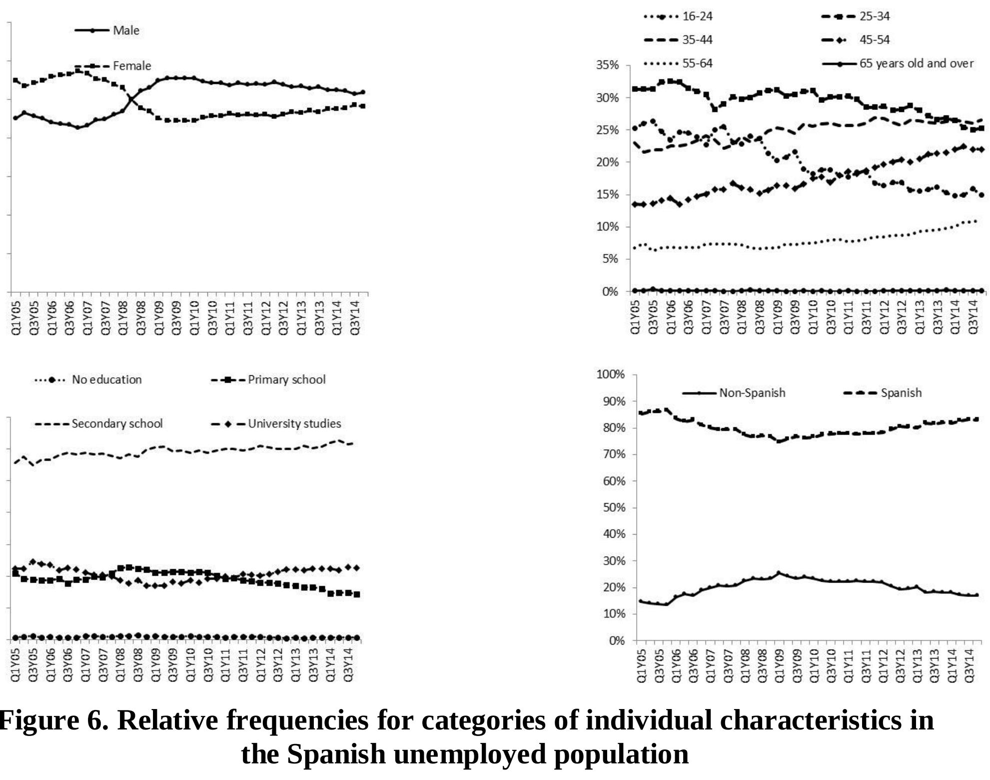


**Figure 6. Relative frequencies for categories of individual characteristics in the Spanish unemployed population**



DOI: 10.24818/18423264/52.1.18.10

### 3. Forecasts of unemployed populations in the four quarters in 2015 and 2016

The unemployed people in each quarter in 2014 are considered to be the original population to which the genetic algorithm is applied to forecast the composition of the unemployed population in the same quarter in 2015 and 2016. To define the selection operator, the forecasts of the unemployment rates by structure, $\hat{f}_{i,k,T+h}$, in the four quarters in 2015 and 2016 need to be obtained from the quarterly rates observed from 2005 to 2014. However, if structure $E_i$ is not present in the active population in quarter$k$ ofyear $t$, then the corresponding $f_{i,k,t}$ is not observed. In these cases, to estimate models as in Equations (2.a) and (2.b), missing data in the series $\{f_{i,k,t}\}_{t=2005,\ldots,2014}$ are replaced by an average of such rates in the nearest neighbors or contiguous quarters. Then, by applying the selection operator defined in section 2, an intermediate population is generated, which is composed of 500,000 individuals to facilitate the calculation of the relative weights of the different structures. This is also the number of individuals in the final population. The intermediate population obtained from the selection is transformed by applying the transition matrix $M$, defined by assuming that the non-transition probabilities are equal to 0.99.

#### 3.1. Results

The relative weights of the different structures in the unemployed population for original, forecasted and observed populations in each quarterare not shown to save space. However, in Figures 7 and 8 the relative weights for all the structures in the final population obtained by genetic algorithms in each quarter in 2015 and 2016 are compared with the observed ones in the same quarter.As observed in these figures, the genetic algorithm provides accurate forecasts of the relative weights by profile in each quarter in 2015 and 2016. In Tables2.a and 2.b, the calculations of the sum of squared errors in these forecasts by profile are shown and compared with the observed ones for a naive forecast obtained by assuming the relative weights observed in 2014 to be the same in the forecasted population. Finally, in Figures 9.a to 9.d and 10.a to 10.d, the forecasted relative weights for the categories of each of the individual attributes (gender, age, education level and nationality) are compared with the observed ones in each one of the quarters in 2015 and 2016.



DOI: 10.24818/18423264/52.1.18.10

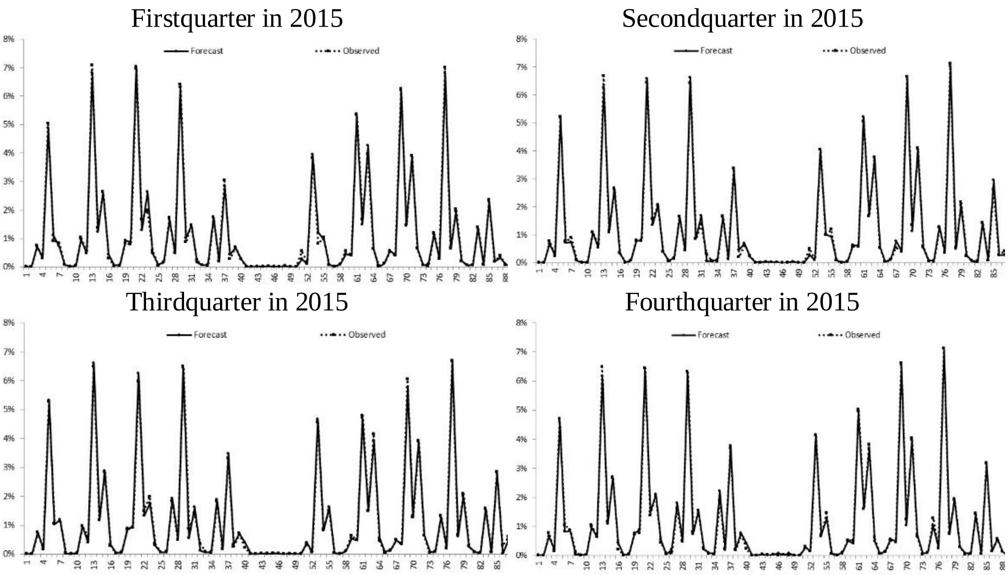


**Figure 7. Forecasted and observed relative weights for all the structures in the four quarters in 2015.**

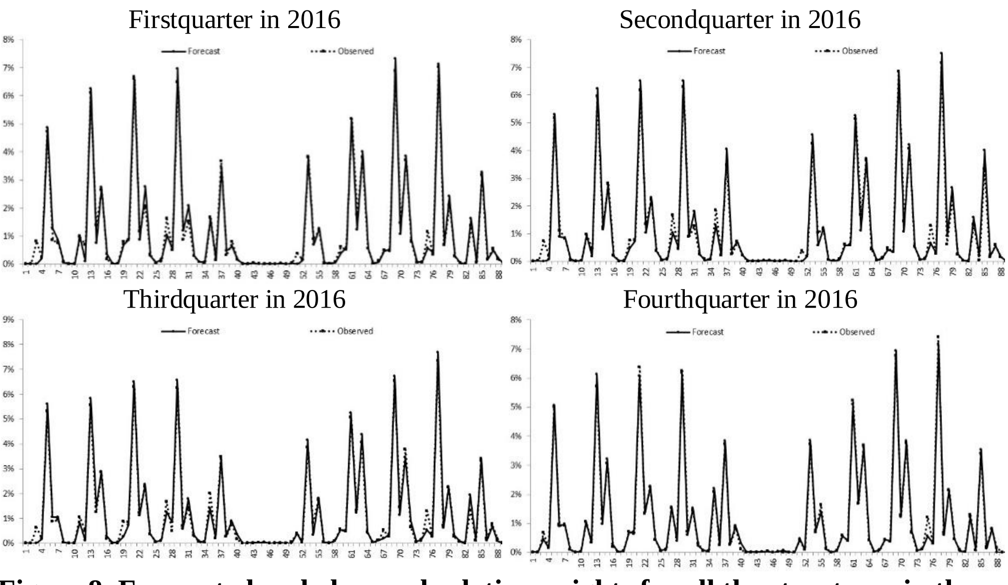


**Figure 8. Forecasted and observed relative weights for all the structures in the four quarters in 2016.**



DOI: 10.24818/18423264/52.1.18.10

**Table 2.a. Sum of non-weighted squared errors in the forecast by profile in the four quarters in 2015 and 2016.**

| Year | 2015 | | 2016 | |
|---|---|---|---|---|
| Quarter | GA forecast(1) | Naive forecast(2) | GA forecast(1) | Naive forecast(2) |
| Q1 | 0.000134598 | 0.000291049 | 0.000484409 | 0.000693736 |
| Q2 | 0.000101297 | 0.000322359 | 0.000479139 | 0.000535125 |
| Q3 | 0.000079680 | 0.000291941 | 0.000419936 | 0.000502013 |
| Q4 | 0.000076181 | 0.000299521 | 0.000119501 | 0.000540067 |

(1)Forecasts obtained by using genetic algorithms.
(2)Forecasts calculated as the relative weights observed in 2014.

**Table 2.b. Sum of weighted squared errors in the forecast by profile in the four quarters in 2015 and 2016.**(1)

| Year | 2015 | | 2016 | |
|---|---|---|---|---|
| Quarter | GA forecast(2) | Naive forecast(3) | GA forecast(2) | Naive forecast(3) |
| Q1 | 0.000002977 | 0.000011131 | 0.000008881 | 0.000029724 |
| Q2 | 0.000002584 | 0.000013896 | 0.000009952 | 0.000024677 |
| Q3 | 0.000001868 | 0.000011902 | 0.000007920 | 0.000020533 |
| Q4 | 0.000001591 | 0.000011097 | 0.000003479 | 0.000023429 |

(1)Squared errors by profile are weighted by the relative frequency observed in the forecasted period.
(2)Forecasts obtained by using genetic algorithms.
(3)Forecasts calculated as the relative weights observed in 2014.

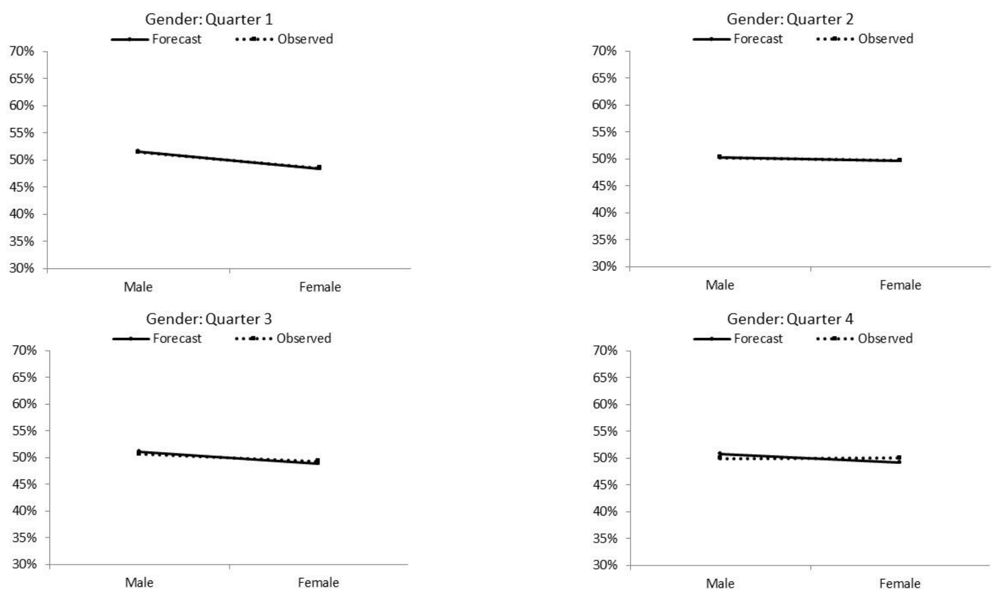


**Figure 9.a. Forecasted and observed relative weights according to gender in 2015.**



DOI: 10.24818/18423264/52.1.18.10

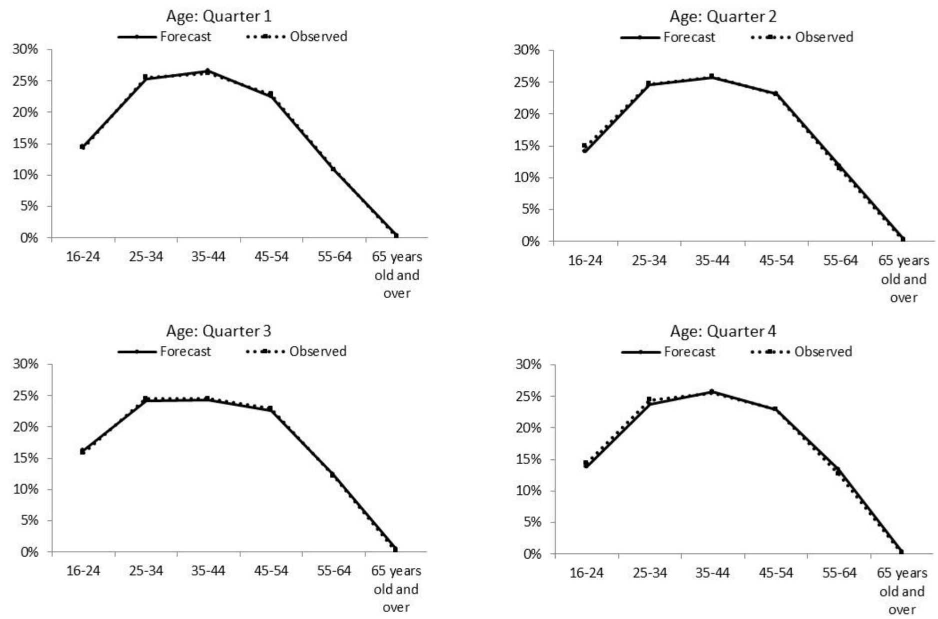


**Figure 9.b. Forecasted and observed relative weights according to age in 2015.**

### 3.2. Discussion

As regards gender, the genetic algorithm under-forecasts the relative weight for females in the Spanish unemployed population in some quarters (Figures 9.a and 10.a). As Peña-Boquete (2014) pointed out, the gap in unemployment rates between males and females had almost disappeared during the economic crisis. However, as shown in Figure 1, women in Spain are again starting to suffer higher unemployment rates than men during the last few years. As predicted by De la Rica and Rebollo-Sanz (2016), gender convergence is only expected to persist in long-term unemployment rates. Furthermore, as observed by Razzu and Singleton (2016) in the United States, the countercyclical flow rate from inactivity to employment can be more intense for women. Figure 4 shows that the increase in female participation in the active population has not stopped yet. Therefore, female participation in the unemployed population is also growing. In fact, the genetic algorithm seems to be capturing such a trend, because the forecasts for this female participation have also increased from 48.4% for the first quarter in 2015 to 49.2% for the last quarter in 2015 and to 49.8% for the last quarter in 2016.



DOI: 10.24818/18423264/52.1.18.10

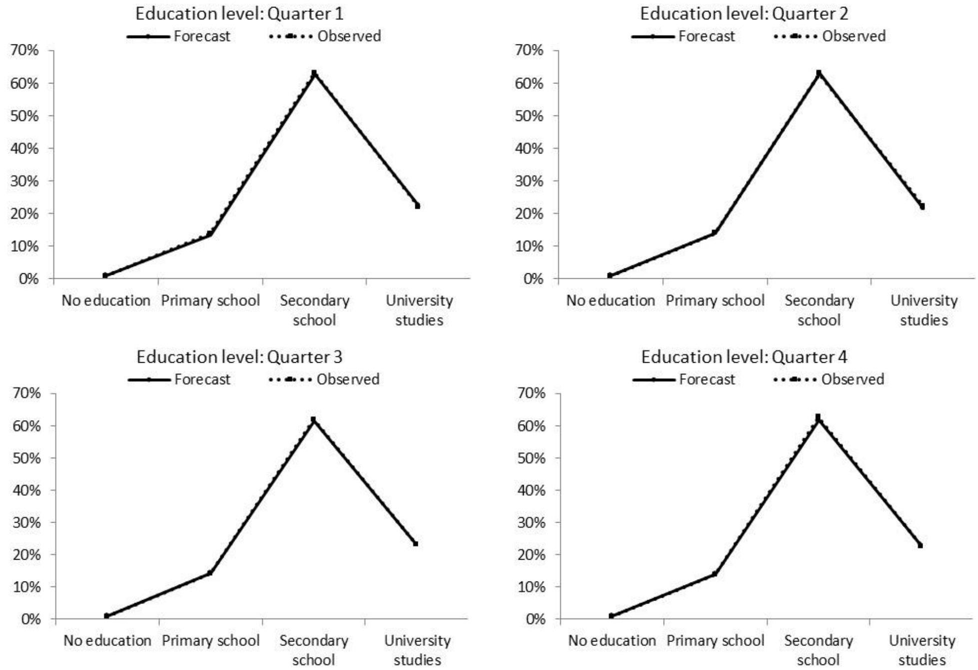


**Figure 9.c. Forecasted and observed relative weights according to education level in 2015.**

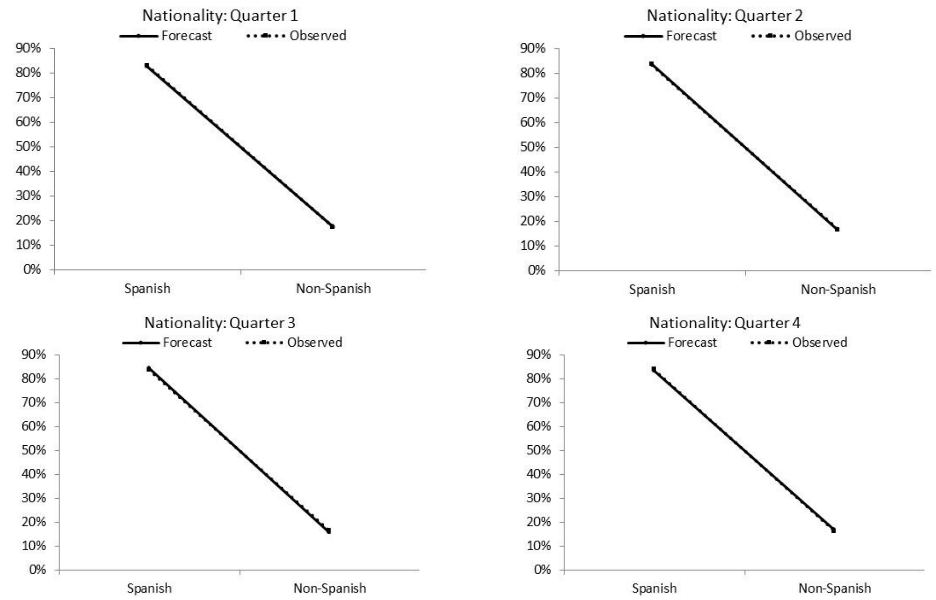


**Figure 9.d. Forecasted and observed relative weights according to nationality in 2015.**



DOI: 10.24818/18423264/52.1.18.10

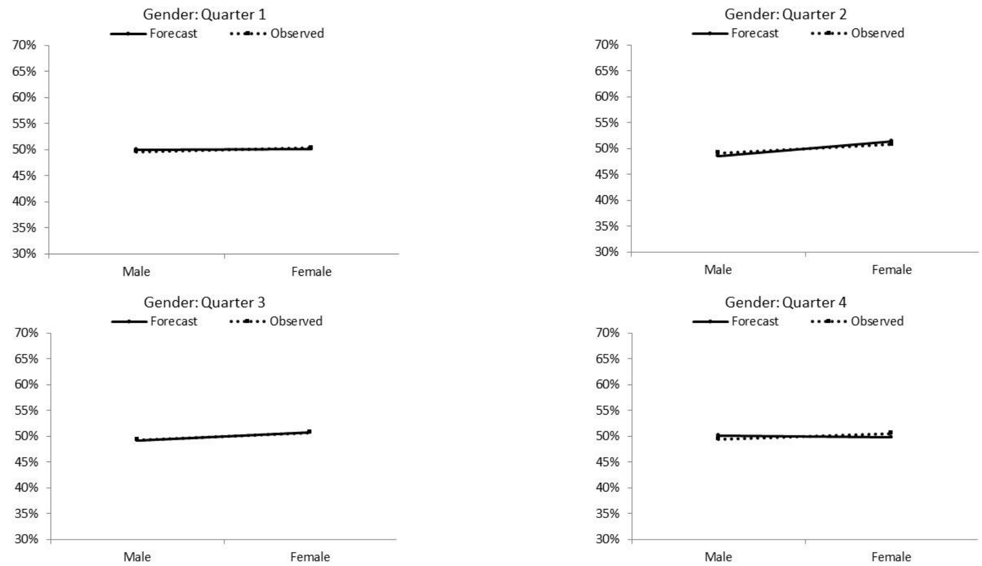


**Figure 10.a. Forecasted and observed relative weights according to gender in 2016.**

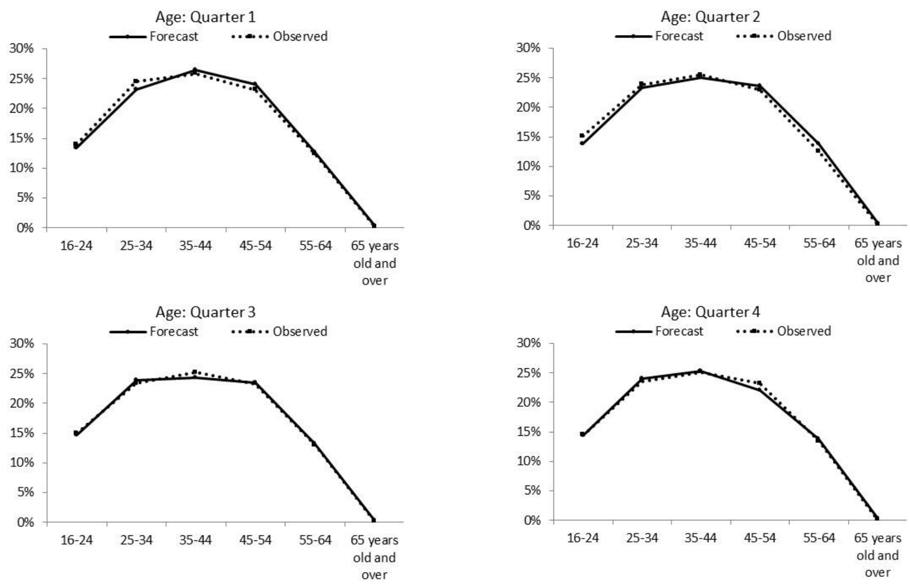


**Figure 10.b. Forecasted and observed relative weights according to age in 2016.**

The forecasts for age segments are more accurate (Figures 9.b and 10.b). The highest unemployment rates for young people observed in Figure 1 reveal that



DOI: 10.24818/18423264/52.1.18.10

the so called Spanish puzzle by Sacht (2015) has not been solved. However, the decrease in young people in the active population, as shown in Figure 4, explains why the first two age segments maintain their relative frequency in the unemployed population.

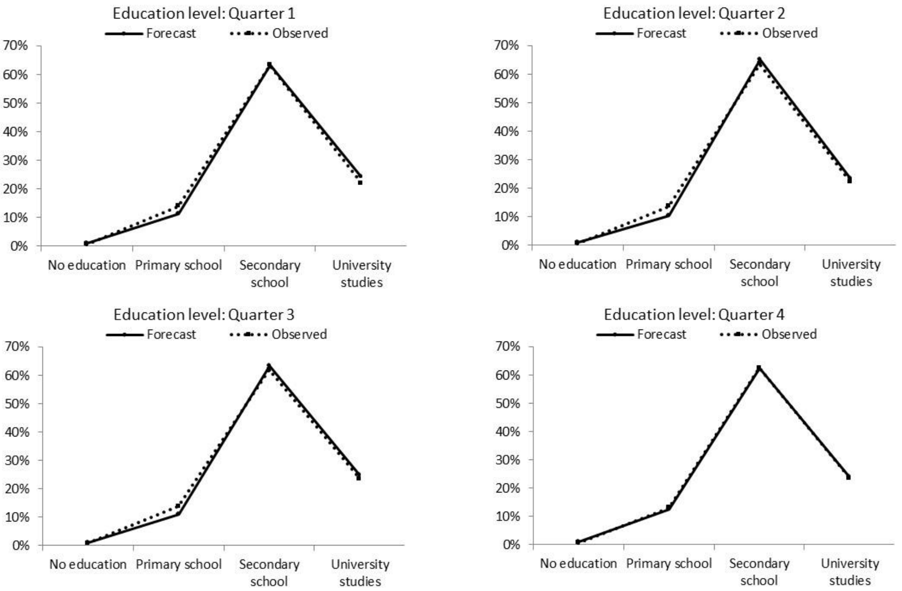


**Figure 10.c. Forecasted and observed relative weights according to education level in 2016.**

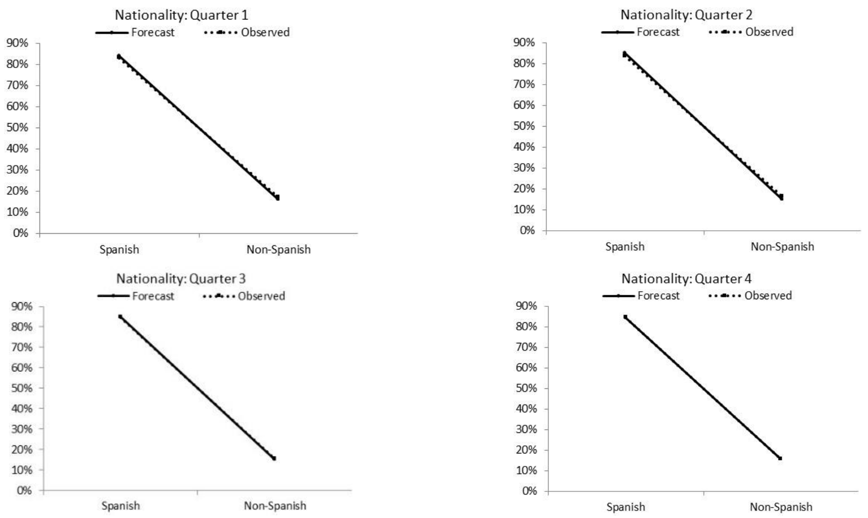


**Figure 10.d. Forecasted and observed relative weights according to nationality in 2016.**



DOI: 10.24818/18423264/52.1.18.10

Figures 9.c and 10.c show the forecasts for educational level. As explained by Klein (2015), the gap in unemployment rates between the low-educated and medium and highly educated people is substantially widened when macroeconomic conditions become worse. In spite of the highest unemployment rates for the lowest educational levels (Figure 1), there is a decrease in their participation in the active population (Figure 4). The genetic algorithm also seems to capture a decrease in their participation in the unemployed population. In addition, the genetic algorithm also provides an accurate forecast of the increase in the participation of university graduates in the unemployed population, which is explained by the increase in their presence in the active population in spite of their lower unemployment rates.

The forecasts for the relative weight for nationalities in the unemployed population are shown in Figures 9.d and 10.d. Except for first and fourth quarters in 2015, the genetic algorithm tends to under-forecast the relative weight for migrants. As mentioned by Motellón and López (2014) and Gil-Alonso and Vidal-Coso (2015), immigrant workers have experienced higher rates of job loss during the Spanish economic recession. However, in the last few years the unemployment rate for migrants is decreasing (Figure 1), and their participation in the active population has also been decreasing since the beginning of the economic crisis (Figure 4). Therefore, the participation of migrants in the unemployed population is diminishing, in accordance with the genetic algorithm forecast for this group, which decreased from 17.6% for the first quarter in 2015 to 16.7% for the last quarter in 2015 and to 15.5% for the last quarter in 2016.

## 4. Conclusions

The genetic algorithm proposed in this paper forecasts the relative weights corresponding to different profiles in the population of unemployed people in Spain. Note that the population forecasted each time the genetic algorithm is performed may differ, but the forecast of the relative weight corresponding to a profile is unlikely to be far from its corresponding final probability, i.e., its expected relative weight.

A critical point in this methodology is the accuracy in forecasting both unemployment rates and active population by profile. Furthermore, the introduction of new characteristics in the profile of unemployed people would require a new design of the algorithm in order to achieve good forecasting performance. Of course, generalization of this methodology to a new context is an empirical matter subject to the evaluation of readers. Nevertheless, the proposed methodology is flexible enough to adapt to more general settings.

Finally, it should be noted that the specific results obtained in this paper are not its main contribution to the literature; rather it is the proposal of a methodology able to provide the type of forecast that could be very useful for designing policies to promote access to the labor market for the unemployed.



DOI: 10.24818/18423264/52.1.18.10

## REFERENCES

[1] **Altuzarra, A. (2015),** ***Measuring Unemployment Persistence by Age and Gender***. *Romanian Journal of Economic Forecasting*, 18(4), 110-133;
[2] **Antón, J-I., Muñoz, R. and Carrera, M. (2012),*****Raining Stones? Female Inmigrants in the Spanish Labour Market***. *Estudios de Economía*, 39(1), 53-86;
[3] **Caliendo, M. and Schmidl, R. (2016),*****Youth Unemployment and Active Labor Market Policies in Europe***. *IZA Journal of Labor Policy,* 5, 1-30;
[4] **Cuestas, J.C. and Gil-Alana, L.A. (2017),*****Unemployment Rate Cycles in Europe***. *Applied Economic Letters*, 24(2), 136-139;
[5] **Dustmann, C., Glitz, A. and Vogel, T. (2010),*****Employment, Wages, and the Economic Cycle: Differences between Immigrants and Natives***. *European Economic Review,* 54, 1-17;
[6] **García-Cintado, A., Romero-Ávila, D. andUsabiaga, C. (2014),*****Spanish Regional Unemployment***. Springer Briefs in Economics. London: *Springer;*
[7] **Gil-Alonso, F. and Vidal-Coso, E. (2015),*****Foreign Immigrants in the Spanish Labour Market: Are They More Resilient or More Vulnerable to the Economic Crisis?*** *Migraciones*, 37, 97-123;
[8] **Ghosray, A., Ordoñez, J. and Sala, H. (2016),** ***Euro, Crisis and Unemployment: Youth Patterns, Youth Policies?*** *Economic Modelling*, 58, 442-453;
[9] **Hernández-López, M. and Cáceres-Hernández, J.J. (2007),*****Forecasting Tourists' Characteristics by a Genetic Algorithm with a Transition Matrix***. *Tourism Management*, 28, 290-297;
[10] **Hernández-López, M. and Cáceres-Hernández, J.J. (2016),*****Forecasting the Composition of Demand for Higher Education Degrees by Genetic Algorithms***. *Economic Computation and Economic Cybernetics Studies and Research*; *ASE Publishing;* 50(3), 153-172;
[11] **Klein, M. (2015),*****The Increasing Unemployment Gap between the Low and High Educated in West Germany. Structural or Cyclical Crowding-Our?*** *Social Science Research*, 50, 110-125;
[12] **Li, T. and Zhang, J. (2010),*****What Determines Employment Opportunity for College Graduates in China after Higher Education Reform?*** *China Economic Review*, 21, 38-50;
[13] **Marelli, E. and Vakulenko, E. (2016),*****Youth Unemployment in Italy and Russia: Aggregate Trends and Individual Determinants***. *The Economic and Labour Relations Review*, 27(3), 387-405;
[14] **Marqués, A.I., García, V. and Sánchez, J.S. (2013),*****A Literature Review on the Application of Evolutionary Computing to Credit Scoring***. *Journal of the Operational Research Society*, 64: 1384-1399;



DOI: 10.24818/18423264/52.1.18.10

[15] **Motellón, E. andLópez-Bazo, E. (2014),*Job Loss among Immigrant and Native Workers: Evidence from Spain's Economic Downturn***. *Social Indicators Research*, 120, 2, 345-371;
[16] **Olmedo, E. (2014),*Forecasting Spanish Unemployment Using Near Neighbour and Neural Net Techniques***. *Computational Economics*, 43, 2, 183-197;
[17] **Pastor, J.M., Peralta, C. andSoler, Á. (2016), *Higher Education as Modulator of Gender Inequalities: Evidence of Spanish Case***. *Higher Education Policy*, 29, 1, 63-88;
[18] **Peiró, A., Belaire-Franch, J. and Gonzalo, M.T. (2012),*Unemployment, Cycle and Gender***. *Journal of Macroeconomics*, 34, 1167-1175;
[19] **Peña Boquete, Y. (2014),*Have the Economic Crises Reduced the Gender Gap on The Spanish Labour Market?*** *Revue de l'OFCE*, 133, 2, 277-302;
[20] **Perugini, C. and Signorelli, M. (2010),*Youth Labour Market Performance in European Regions***. *Economic Change and Restructuring*, 2, 151-185;
[21] **Poirier, D.J. (1976),*The Econometrics of Structural Change: With Special Emphasis on Spline Functions***. Amsterdam: *North-Holland;*
[22] **Ponomareva, N. & Sheen, J. (2013),*Australian Labor Market Dynamics across the Ages***. *Economic Modelling*, 35, 453-463;
[23] **Queneau, H. and Sen, A. (2012),*On the Structure of US Unemployment Disaggregated by Race, Ethnicity, and Gender***. *Economics Letters*, 117, 91-95;
[24] **Razzu, G. and Singleton, C. (2016),*Gender and the Business Cycle: An Analysis of Labour Markets in the US and UK***. *Journal of Macroeconomics*, 47, 131-146;
[25] **Rica de la, S. &Rebollo-Sanz, Y.F. (2016),*Gender Differentials in Unemployment Ins and Outs during the Great Recession in Spain***. *De Economist*, 165(1), 67-99;
[26] **Sacht, S. (2015),*On the Efficiency of Labor Market Reforms: How to Solve the Spanish Puzzle?*** *Economics: The Open Access*, 9 (2015-44), 1-28;
[27] **Santana Quintero, L.V. and Coello, C.A. (2006),*An introduction to evolutionary computation and some of its applications in economics and finance***. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 2(1): 3-26;
[28] **Uhlendorff, A. and Zimmermann, K.F. (2014),*Unemployment Dynamics among Migrants and Natives***. *Economica*, 81, 348-367;
[29] **Vicente, M.R., López-Menéndez, A.J. and Pérez, R. (2015), *Forecasting Unemployment with Internet Search Data: Does it Help to Improve Predictions when Job Destruction Is Skyrocketing?*** *Technological Forecasting and Social Change*, 92, 132-139;
[30] **Waltman, L. and van Eck, N.J. (2012),*A Mathematical Analysis of the Long-Run Behavior of Genetic Algorithms for Social Modelling***. *Soft Computing*, 16: 1071-1089.



DOI: 10.24818/18423264/52.1.18.10